SEGMENTAL TEMPO ANALYSIS OF PERFORMANCES IN USER-CENTERED EXPERIMENTS IN THE DISTRIBUTED IMMERSIVE PERFORMANCE PROJECT

Elaine Chew[*], *Alexander Sawchuk*[†], *Carley Tanoue*[*], *Roger Zimmermann*[‡]
University of Southern California Viterbi School of Engineering
Integrated Media Systems Center
Los Angeles, CA 90089, USA

ABSTRACT

In this paper we present a quantitative analysis of performer-based experiment data in the Distributed Immersive Performance Project. The experiments explore the effects of auditory latency on musical ensemble and interpretation in order to determine the thresholds for usability. We propose two measures – the segmental tempo difference and the segmental tempo ratio from a baseline performance – as objective quantifiers of performance strategies. Our earlier analyses of qualitative reports showed that the usability threshold lies between 50 and 75 ms. We demonstrate that the proposed analyses of the performance data, captured in MIDI format, lead to results similar to the reports. The tempo difference and tempo scaling across logical segments of the piece show marked increase in variability when the auditory delay is in the range of 50 to 100 ms (for two examples) and 50 to 75 ms (for the most rapid example). The span of the tempo difference and ratio values at latency 150 ms is less than that around the usability thresholds. We surmise that around the threshold, the users attempt numerous performance strategies to compensate for the delay; at latencies far above the threshold, such strategizing fails and the performers revert to more stable practiced norms. These findings indicate that segmental tempo difference and ratio analyses are useful indicators of performance decisions, and that quantitative analysis of performance data may be a viable way of evaluating the psychophysical effects of collaborative performance under various immersive conditions.

1. INTRODUCTION

The Distributed Immersive Performance (DIP) project explores the creation of a seamless environment for remote and synchronous musical collaboration. The goal of the user-centered experiments in the DIP project is to systematically and comprehensively study and assess, both qualitatively and quantitatively, the psychophysical effects of latency on remote collaborative musical performance over the Internet. To further this goal, we enlisted the help of two expert users, the Tosheff Piano Duo [16], to perform a rhythmically demanding twentieth century classical musical piece, Poulenc's *Sonata for Piano Four-Hands*, under different conditions of auditory delay. The movements of the piece range from slow to fast to very fast. We have recorded video, audio and MIDI data streams from these experiments, as well as documented the duo's self-reports on performing under the different conditions. This paper presents an evaluation methodology that focuses on high-level interpretative choices to compare performances of the same piece under different immersive conditions. We apply the method to the MIDI data captured in the DIP experiments to show that the users' self-reported thresholds for usability are reflected in the MIDI data under these objective measures.

The proposed evaluation method considers segments of the piece based on motivic combination, and compares the tempo difference and tempo scaling of these logical segments against a baseline performance. It assumes that the performers have a preferred tempo at which to traverse each segment, which can be inferred from a baseline performance. The tempi of these segments in other performances are then compared to the baseline figures. Prior analyses of the duo's answers to a questionnaire have indicated that they found the usability threshold for auditory delay to be in the range of 50 to 75ms. Analysis of the segmental tempo difference and segmental tempo scaling data shows a marked increase in variability of these measures in the range of 50 to 75ms (for one of the movements of the Poulenc piece) and 50 to 100ms (in the other two movements), which correspond to the performers' self reports on usability. Hence, segmental tempo difference and scaling analyses show promise as quantitative and objective methods for revealing usability thresholds in remote collaborative performance.

---

[*] Epstein Department of Industrial and Systems Engineering
[†] Department of Electrical Engineering – Systems
[‡] Department of Computer Science

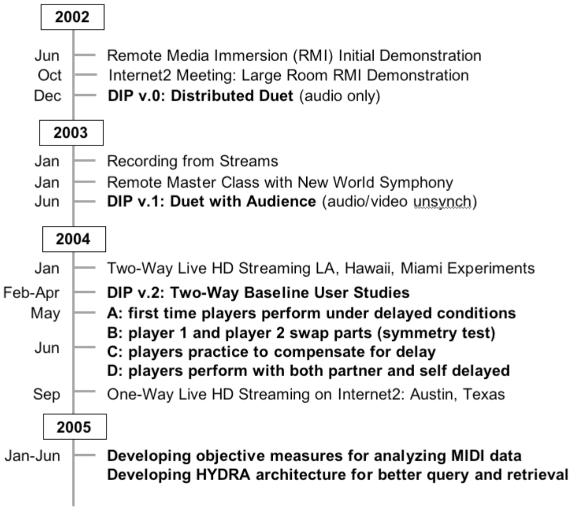
We surmise, based on the increased variability away from the baseline, that at the threshold boundary, the duo increases their rate of attempting other interpretative decision strategies that vary from their norm to compensate for the auditory delay. There is low tempo variability before the threshold because the auditory delay is deemed to be acceptable for their usual mode of playing. Beyond the threshold, the variability is also reduced. We hypothesize that at this point, the tempo is considered unacceptable for effective collaborative performance, and the duo revert to their modus operandi, and simply proceed with a more stable strategy closer to their practiced norm in the absence of adequate auditory feedback conditions.

The findings of our systematic latency studies in the DIP project have ramifications for the design of remote collaborative systems at large, and impact any application in which users perform a synchronous and coordinated task. Musical collaboration, a temporally exacting activity demanding multiple levels of intercommunication among the players, serves as an ideal testbed for studying the system requirements for synchronous collaborative activities over distance. The rest of this introductory section presents a brief history of our DIP experiments and some related work. The remainder of the paper is structured as follows: Section 2 describes the physical and technical setup of the psychophysical experiments, as well as the experimental design. Section 3 reviews the users' self-reported assessments in each set of experiments. Section 4 formally describes the analysis method. Section 5 presents the results when applied to the experimental data, and a discussion of the findings. Finally, Section 6 follows with conclusions.

1.1. History of the DIP Experiments

The psychophysical experiments in DIP are the latest in a succession of experiments in distributed performances [11] at the Integrated Media Systems Center (IMSC) [7]. In this section, we present a brief overview of the experiments that have motivated our design of a systematic course of evaluation of the effects of latency on remote collaboration over the Internet. Table 1 shows a timeline of our DIP and related experiments involving Internet streaming of musical performances.

The first DIP experiment involving synchronous collaboration between two connected sites took place in December of 2002, an informal test of the viability of distributed performance with auditory delay. The two musicians, Wilson Hsieh (viola) and Elaine Chew (piano), were located in the Electrical Engineering Building (EEB) and Powell Hall (PHE) respectively. They were connected by two-way audio, with immersive audio in EEB and stereo audio in PHE. They played Piazzolla's *Le Grand Tango,* and excerpts from Hindemith's *Sonata Op. 11 No. 4.* Chris Kyriakakis, using a ProTools console, artificially introduced audio delays that ranged from 20 to 300 ms. He also used 10.2-channel immersive audio [10] to simulate the acoustics of a concert hall at the violist's site. The violist found that the use of 10.2-channel immersive audio greatly ameliorated the discomfort of playing in the distributed environment. Another observation was the difference in perspectives at the two sites: if the signals from EEB are labelled A and those from PHE are labelled B, then EEB experiences {A, B+d }, where d is the delay added to the signal in transit, and the people located at PHE experience { A+d, B }.



2002	
Jun	Remote Media Immersion (RMI) Initial Demonstration
Oct	Internet2 Meeting: Large Room RMI Demonstration
Dec	**DIP v.0: Distributed Duet** (audio only)
2003	
Jan	Recording from Streams
Jan	Remote Master Class with New World Symphony
Jun	**DIP v.1: Duet with Audience** (audio/video unsynch)
2004	
Jan	Two-Way Live HD Streaming LA, Hawaii, Miami Experiments
Feb-Apr	**DIP v.2: Two-Way Baseline User Studies**
May	**A: first time players perform under delayed conditions**
	B: player 1 and player 2 swap parts (symmetry test)
Jun	**C: players practice to compensate for delay**
	D: players perform with both partner and self delayed
Sep	One-Way Live HD Streaming on Internet2: Austin, Texas
2005	
Jan-Jun	**Developing objective measures for analyzing MIDI data**
	Developing HYDRA architecture for better query and retrieval

Table 1. Timeline of DIP and related experiments.

The second DIP experiment, labeled DIP v.1 in the timeline in Table 1, took place in June 2003. The pianist, Elaine Chew, was located in Ramo Hall, while keyboard faculty Dennis Thurmond, on an accordion, was co-located with the audience in PHE. The two sites were approximately 350m apart on the USC campus. The audience site was equipped with 10.2-channel sound while the pianist heard the performance at PHE through an ear piece. Both sites were connected by video; no effort was made to synchronize the video and the sound. Photos from the two sites in the experiments are shown in Figure 1.

Figure 1. DIP v.1 with USC keyboard faculty, Dennis Thurmond, in PHE 106 (left), and Elaine Chew in Ramo Hall 106 (right), located approximately 350m apart on the USC campus.

Because the audience was co-located with one of the musicians (Thurmond), the performance was constrained so that the end-result was as convincing as possible at the audience-site. The video delay was also found to be too large to be usable as a means of giving cues for synchronization, and the musicians relied only on the audio to create the performance. This arrangement resulted in performances that were more musically restrained than the performers would have preferred.

Our prior experience with distributed performance showed that there was much room for improvement in the design of systems for remote collaboration. Hence, the need exists to design experiments to identify the important system parameters for minimum usability, such as the maximum tolerable audio delay, the maximum tolerable audio/video (combined) delay, minimum display size, appropriate viewing direction, inclusion of multi-channel audio, etc. Many of these system parameters for minimum usability are unknown due to the lack of sufficient previous studies in distributed live music performances. To address this gap in knowledge about remote collaborative systems, we conducted several sets of experiments in May and July 2004 to systematically document and study the effects of auditory delay on performers' satisfaction with the ease of creating a tight ensemble, a musical interpretation, and adaptation to the conditions. Results from the experiments will shed light on guidelines for designing economically viable remote and synchronous collaborative systems.

1.2. Related Work

Other research groups have experimented with distributed collaborative performance environments.

The experiments closest in spirit to ours are those on the defining of delay thresholds for clapping synchronization, a special case of musical ensemble, conducted at Stanford's SoundWIRE Group [15] by Chafe et al [1] and Shuett [13]. Two players were located in separate rooms (isolated both visually and aurally), and latency was artificially added to each of their auditory signals from source to listener. Each take was recorded and an onset detection algorithm employed to determine clapping discrepancies. The subjects were asked to clap a simple rhythm at designated tempi, with different ensemble strategies and delay policies. In Shuett's [13] pilot study, the results were evaluated using two measures – tempo direction and tempo deviation – with a proposed third synchrony measure. The study showed that the tempo would slow down for delays greater than 30 ms, that the delay threshold was between 50 and 70 ms if one subject was designated the leader, and delays of 10 to 20 ms had a stabilizing influence on the tempo. Chafe et al [1] used automatic methods for event detection and tempo tracking, and measured tempo consistency as a function of delay. This study found that longer delays produced increasingly severe tempo slow downs, while shorter delays (<11.5 ms) produced surprisingly some tempo acceleration, and that moderate amounts of delay are beneficial to tempo stability.

A number of distributed performance events have transpired over the years with the development of the Internet and streaming technologies. We list here a few examples. In 1993, Eve M. Schooler et al [12] demonstrated a network "Flow Synchronization Protocol" in a distributed trio performance, with piano (designated conductor) at the University of Southern California (USC) in Los Angeles, and violin and cello in Boston, playing Haydn's *Piano Trio, No. 1 in G*, Finale. At other performances, including at ACM Multimedia in San Francisco (Oct 1994), the group created and synchronized three real-time streams of music from different Internet hosts with delays in the order of 200 ms. An audience was scattered across DARTnet with the funding sponsor listening in DC. One conclusion was that it was difficult for performers to be listeners, which is counter to typical performance practice.

The SoundWire group from Stanford University [15] has developed technologies for low-latency audio streaming [2] and performed many distributed collaborative experiments involving a cellist playing an electric cello from various locations around the world and a pianist stationed in a studio at CCRMA. The piece performed was Brahms' *Sonata for Piano and Violoncello in E minor, Op. 38*. At a successful streaming session, the cellist was playing at the Internet 2 Headquarters in Armonk, NY. There was little signal loss and the acoustic latency was "on the 'hairy edge' for an unencumbered performance" [15,6].

Another collaborative performance between McGill and Stanford, held on June 13, 2002, used McGill's low-latency ultra-videoconferencing system and next generation research and education networks, CA*net 3 and Internet 2 [6]. The event featured full-screen bidirectional video and multi-channel audio; the audio latency was reduced when the video was turned off.

The GigaPop Ritual [9] was a 2003 distributed collaborative performance between McGill and Princeton University at NIME 2003. The performers played an original piece titled "A Live Networked Performance Piece for Two Electronic Dholaks, Digital Spoon, DigitalDoo, 6 String Electric Violin, Rbow, Sitar, Tabla, and Bass Guitar." The performance involved hi-bandwidth, bi-directional real-time streaming of audio, video, and controller data from multiple sources, but did not explore the effects of latency on the musicians. Instead, the musicians explored different rhythms and soundscapes, reacting in free-form improvisation to one another over a high-bandwidth network.

Related networked performances include one by Scot Gresham-Lancaster et al's that occurred between

Vancouver, Marseilles, and Troy at NIME 2005 utilizing AB_Time [8], and a one-way performance between McGill and USC in Los Angeles [3] at the 109th Audio Engineering Society's Convention. At the AB_Time performance, the sounds of each space were mapped into the other spaces; there was considerable signal latency and interaction was restricted to free-form improvisation by the players and dancers. At the AES convention, the demonstration presented the Internet transmission of multichannel music in high-resolution 24bit/96kHz PCM and MPEG AAC with video and spatialization control between McGill and USC, and was a first in terms of demonstrating the feasibility of compelling possibilities for remote audio mixing applications. Other one-way streaming of high definition video and multi-channel audio experiments have been led by the team at USC’s IMSC, including the streaming of classical performances from the New World Symphony to USC in 2003, and between two concert halls at the University of Texas, Austin, in 2004 (see timeline in Table 1).

To the best of our knowledge, our experiments involving professional musicians performing complex composed pieces is the first evaluation experiments on such a realistic scale.

2. DIP v.2 EXPERIMENTS

This section describes the experiment and system design for our performer-centered experiments in DIP v.2.

2.1. Expert Users and Physical Setup

Our musical collaborators are the Tosheff Piano Duo [16], an award-winning professional piano duo comprising of Vely Stoyanova and Ilia Tosheff. By engaging professional players in the study, we eliminate, or at least minimize, issues associated with learning and adaptation in duo playing that could bias our results. The Tosheff Duo have been performing together since 1997, and arrived at the experiments having practiced the assigned piece.

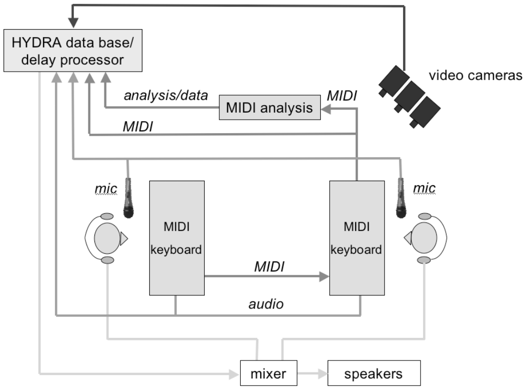


Figure 2. DIP v.2 physical setup.

In the DIP v.2 experiments, we seat the two musicians in the same room across from each other so that the visual contact is immediate, and visual delay is effectively zero. The physical setup of the experiments is shown diagrammatically in Figure 2. Each musician plays on a Yamaha P80 MIDI keyboard with 88 weighted-action keys. MIDI events, audio output, and video captured on three HD cameras are streamed concurrently to the High-performance Data Recording architecture (HYDRA) [17,18,19] database, which will be described further in the next section. An audio delay box (a Protools console) and low latency audio streaming, developed by Christos Papadopoulos and Rishi Sinha, regulate the delay in each audio stream.

2.2. Recording and Playback

The recording, archiving and playback of performances requires a multi-channel, multi-modal recording system that can store a distributed performance event in real-time and be capable of playing back the event with user defined delay offsets between the various streams that constitute the performance. The HYDRA system enables the real-time recording and documentation, and the distribution of multi-modal information, which facilitates comprehensive analysis and evaluation of the psychophysical effects of latency and fidelity on music and other forms of human interaction among interconnected sites.

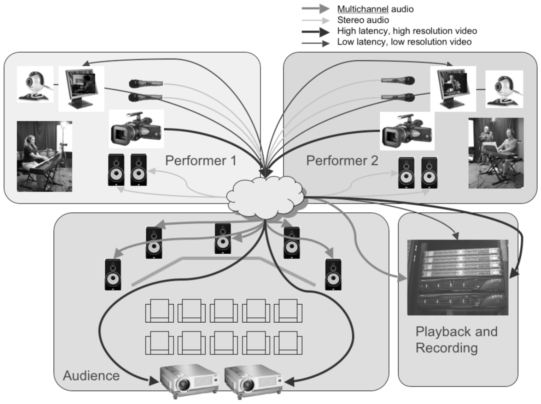


Figure 3. DIP v.2 technical setup. Thin/thick lines denote low /high bandwidth signals, respectively.

Figure 3 shows the technical setup for DIP v.2, and the relation of the HYDRA architecture to the various data streams captured in real time. HYDRA is an extension of an earlier streaming media system called YIMA [14,20], and incorporates real-time stream recording. In the experiments, two digital video streams were acquired from JVC JY-HD10U cameras via FireWire (IEEE 1394) in HDV format (1280x720 pixels at 30 frames per second). The resulting MPEG transport streams were time-stamped and transmitted in discrete packets to the storage backend at approximately 20 Mb/s over an IP network. The front-end and back-end dialog includes control commands such as record, pause, resume, and stop. For monitoring purposes both camera streams were rendered with a software decoder on the HYDRA front-end machine. Two uncompressed PCM audio channels were acquired (16-bits per sample at 48 KHz sampling rate). The bandwidth required was approximately 1.5 Mb/s. The MIDI data produced by the electronic pianos were time-stamped and forwarded to the back-end where they were stored as discrete events in a database.

2.3. Experimental Design

The test piece that the duo was assigned to practice was Poulenc's *Sonata for Piano Four-Hands*, a rhythmically challenging twentieth century classical composition. The duo was asked to perform each movement of the sonata under different conditions of auditory delay. The three movements of the Poulenc piece ranged from fast, to slow, to very fast. Their attributes are summarized in Table 2.

Movement	Notated tempo (bpm)	Meter	Tatum (note value grid)	Reqd IOI (ms)
I.Prelude	132	4/4	8th-note	227
II.Rustique	46	4/4	8th-note	652
III.Final	160	4/4	16th-note	94

Table 2. Movement attributes in Poulenc's Sonata.

Experiment A: The players are asked to perform a movement of Poulenc's Sonata under a randomly assigned level of auditory delay unknown to them. The schematic description of the experimental setup is shown in Figure 4.

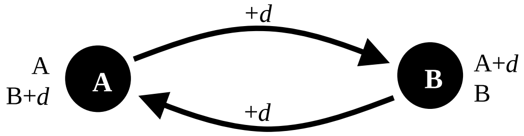


Figure 4. Auditory delay setup for experiments A, B, and C.

The two players are given approximately 30 seconds to calibrate to the given conditions by playing, in unison, the Seconda part in the first section of the first movement (a rhythmic pattern repeated numerous times). They are then asked to perform the movement as best they can under the conditions. The audio latency ranged from 0 ms to 150 ms. Their musical interactions and performance nuances are recorded as HD video, audio and MIDI streams with common time stamps.

At the end of the movement, the players complete a questionnaire. The questionnaire consisted of the following questions (to be answered on a scale of 1 to 7, with 1 being the easiest and 7 the most difficult):

- *(a) How would you rate the ease of ensemble playing?*
- *(b) How would you rate the ease of creating a musical interpretation?*
- *(c) How would you rate the ease of adapting to this condition with practice?*

After completing the questionnaire, the duo is asked to self-report on their observations in a short debriefing session.

To ensure that the experiments were not biased by the players' individual personalities and performance tendencies, we devised the symmetric experiment B.

Experiment B: Same as Experiment A, except the players swap parts; Prima now played Seconda and Seconda now played Prima.

Experiment C: Same as Experiment A, except the users were asked to practice and strategize to compensate for the delay. Because the delay tolerance threshold appeared to be around 50 ms, experimental sets C and D focused on auditory delay levels in the region around 50 ms, and ranged from 40 to 100 ms.

Experiment D: Same as Experiment A, except delay is added to both player's audio input, as in the case, for example, of hearing the performance from a third perspective. The auditory delay setup for this set of experiments is shown in Figure 5. Experimental set D also focused on the region around 50 ms.

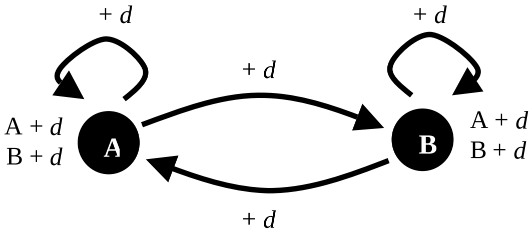


Figure 5. Auditory delay setup for experiment D.

For experimental sets A, C, and D, Vely Stoyanova played the Prima part and Ilia Tosheff played the Seconda part. For experimental set B, Vely played the Seconda part and Ilia the Prima part.

3. PRIOR RESULTS

In this section, we summarize the users' questionnaire answers and qualitative self-reports on usability under different auditory delay conditions. The detailed report of these experiments can be found in [4,5]. The users' assessments of usability thresholds are given by the results from Experiments A & B, and compensating strategies are given by the qualitative assessments in Experiments C & D.

3.1. Delay Tolerance Threshold

As reported in [4], the users' responses to the three questions for Experiment A as well as the results of the symmetry experiment (B) showed that the duo found delays under 50 ms to be acceptable, and struggled to keep time when the delay rose above 50 ms. The results are summarized in Figure 6. In general, the duo felt that musical interpretation was compromised when the delay was above 50 ms. At around 50 ms, they were conscious of the delay but felt that they might be able to compensate with practice. At around 75 ms, musical ensemble was difficult. However, both players appeared confident that adaptation was possible below 75 ms. At 100 ms, ensemble was extremely difficult; it was almost impossible at 150 ms. For the three movements, the usability threshold appeared to be around 50 ms for both the Prelude and Final, and 75 ms for the Rustique. Latency values beyond these thresholds resulted in difficulty ratings above 3.5.

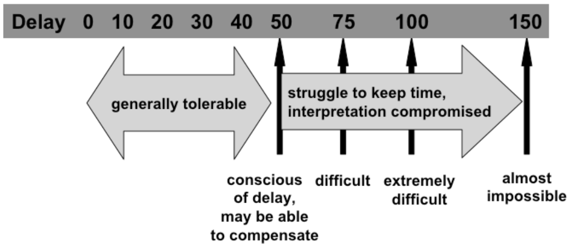


Figure 6. Results summary for Experiments A & B.

In the symmetry experiments B, both players found the unfamiliar part slightly more difficult on all counts.

3.2. When More Delay is Better

In experiment sets C and D, we found that the existence of a common perspective was paramount to the performers' comfort in remote collaboration [5], and that players may be willing to tolerate and adjust to increased delay in feedback of their own actions in order to achieve the experience of a common perspective. The common reality allows for the formulation of joint goals, and strategies for achieving them.

Results for Experiment C: The players devised various strategies to adapt to the trying conditions. The tolerance threshold did not differ from experiments A or B, and remained around 50 ms. Since the experiment set's delay values ranged from 40 to 100 ms (focusing around the tolerance threshold), the frustration and tension between the players steadily increased throughout the experiment.

When each player put on the other person's headphones to better understand their different perspectives, they asked to hear the auditory feed from the perspective of the audience. According to Ilia, when he is playing, he is not thinking about what his hands are doing. He focuses on what it is the audience hears, creates a mental image of what he wishes to portray and lets his hands do the rest. This request for the audience's perspective resulted in experiment set D.

Results for Experiment D: The players' feedbacks of their own playing were delayed to synchronize with the incoming feed from their partner. The duo was noticeably much happier in condition D than in condition C. The overall tolerance threshold, originally at around 50 ms for condition C, was increased to 65 ms under condition D, as shown in Figure 7. The players liked condition D so much that, according to Ilia, the performance could be 'perfect' with practice.

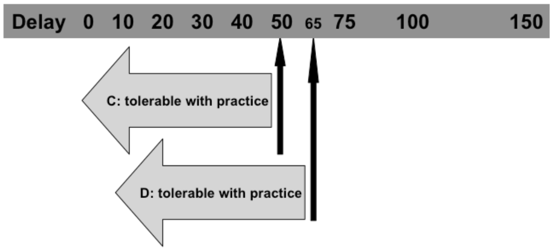


Figure 7. Users' assessment for Experiments C & D.

4. QUANTITATIVE ANALYSIS METHOD

We base our quantitative analysis of the performances on the MIDI stream recorded in the experiments. This section explains formally our method for analyzing the MIDI streams collected in the experiments.

4.1. De-constructing Piece into Segments

The analytical structure of Poulenc's *Sonata for Piano Four-hands* lends itself to segmental analysis. Each movement is constructed from a small set of motivic patterns, and logical segments can be readily identified by their generative patterns, or by tempo change markings indicated by the composer. After segmentation, we measured the length of each segment in each part independently to determine the tempo of that segment.

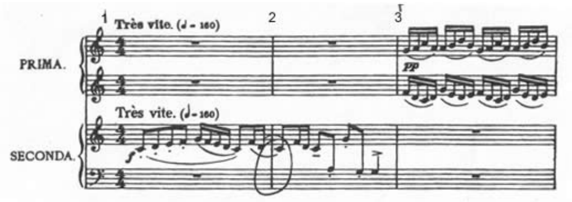



Figure 8. Example of segmentation boundaries: at measure 1 for Seconda and 3 for prima at the beginning of movement 3: Final.

Occasionally, a segment may be silent at the beginning of the assigned boundary beginning only after several bars of silence, as in the Prima part of the beginning of the third movement (see Figure 8). In these cases, the left-most boundary of the segment was re-defined as the onset of the first note in that section. In the case of Figure 8, the first segment for the prima part would begin on bar 3 rather than bar 1.

4.2. Segmental Tempo Difference and Ratio

For each segment, we compared the performed tempo for that segment with a baseline tempo that is obtained from the duo's performance of that same segment under 0 ms delay. We consider two statistics based on segment tempi: (1) the tempo difference between the segment under evaluation and that of the corresponding segment in the baseline performance; and, (2) the tempo ratio between the evaluation and the baseline segment

We first define the segmental tempo difference. Assume that the segment boundary times for a part from the baseline performance is $\{ t_{0,0}, t_{0,1}, t_{0,2}, \ldots, t_{0,S} \}$ (in minutes), where S is the total number of segments. For the same part in a performance with *d* ms auditory delay, let its corresponding segment boundary times be $\{ t'_{d,0}, t'_{d,1}, t'_{d,2}, \ldots, t'_{d,S} \}$. Recall that $t_{d,i}$ is the onset time of the first note in segment *i*. Let the number of beats in segment *i* be b_i. Then, the segmental tempo difference for segment *i* can be written as a function of the auditory delay $(\cdot)$:

$$d_i(\cdot) = \left(\frac{b_i}{t'_{\cdot,i} - t'_{\cdot,i-1}}\right) - \left(\frac{b_i}{t_{\cdot,i} - t_{\cdot,i-1}}\right). \qquad (1)$$

We are also interested in the segmental tempo ratio, the proportional increase or decrease in tempo from the baseline. The segmental tempo ratio for segment *i* can be calculated using the following function:

$$r_i(\cdot) = \frac{t_{\cdot,i} - t_{\cdot,i-1}}{t'_{\cdot,i} - t'_{\cdot,i-1}}. \qquad (2)$$

5. QUANTITATIVE RESULTS

We present the segmental tempo difference and segmental tempo ratio analyses of the DIP experiment data in this section. Table 3 summarizes the segmentation boundaries for both the Prima (p) and Seconda (s) parts in each movement of the sonata, Prelude, Rustique, and Final. Decimal values in the table represent the fractional value within that particular measure (e.g. 22.75 equals the third beat of the 22nd measure in the 4/4 time meter). These segmentation boundaries were then translated to millisecond values, and the length of each segment, and the tempo of that segment is then calculated from these boundary start times.

Sections 5.1 and 5.2 present the analysis of the segmental tempo difference and ratio values for each of the three movements.

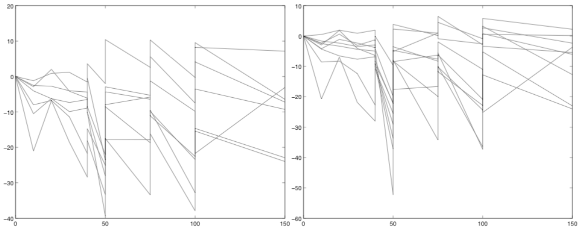
Movement	Part	Segment boundaries (measures)
Prelude	p	{ 3, 17, 26, 37, 45, 61, 70, 71 }
Prelude	s	{ 1, 3, 15, 17, 37, 45, 59, 61, 70, 71 }
Rustique	p	{ 5, 9, 16, 22, 22.6875 }
Rustique	s	{ 1, 5, 9, 16, 22.25, 22.6875 }
Final	p	{ 3, 16, 35, 55, 71, 72 }
Final	s	{ 1, 17, 35, 55, 71, 72 }

Table 3. Segment boundaries for each part (p=prima, s=seconda) of each movement in Poulenc's Sonata.

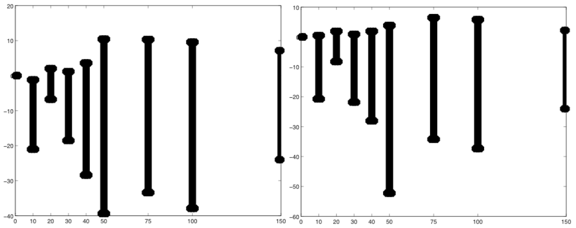
5.1. Segmental Tempo Difference

The charts in Sections 5.1.1 through 5.1.3 present the segment tempo difference analysis results for each of the movements: Prelude (Figure 9), Rustique (Figure 10), and Final (Figure 11). Each figure consists of three parts, (a), (b) and (c). In the (a) figures, the segmental tempo difference (from the baseline) for all segments is plotted against the audio latency. Multiple takes at the same latency are shown as vertical lines.

(a) Segment Tempo Difference vs. Delay



(b) Tempo Difference Span vs. Delay



(c) Magnitude of Tempo Difference Span vs. Delay

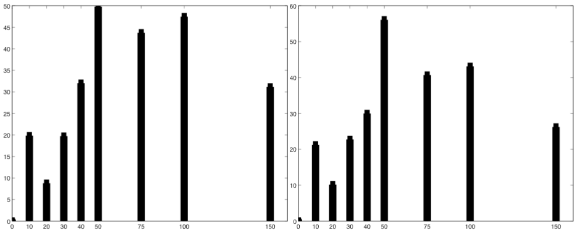


Figure 9. Segmental tempo difference analysis of movement I: Prelude.

The (b) figures highlight the span of the tempo difference, showing both the minimum and maximum values of the tempo difference for each segment, and the (c) figures chart the magnitude of the span of the tempo difference. In each part, two graphs are shown, one for the Prima part (on the left), and one for the Seconda part (on the right).

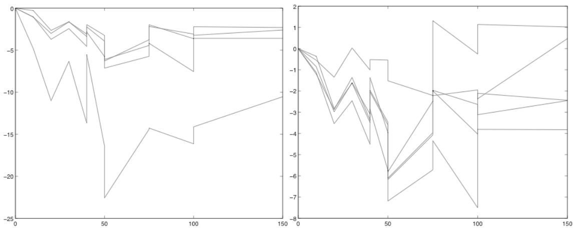
5.1.1. Analysis of Movement I: Prelude

Figure 9 shows the segmental tempo difference from the baseline performance over different auditory delay conditions for performances of the fast and rhythmic Prelude.

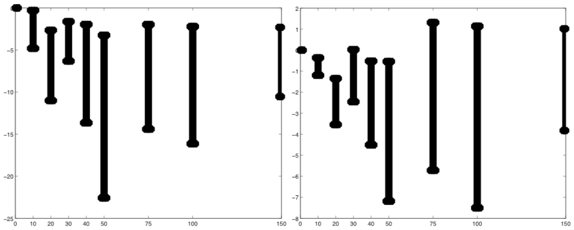
Each line in Figure 9(a) shows a $d_i(\cdot)$ for a given i. The graph shows the collection of plots of $d_i(\cdot)$ for all i. The graphs show the degree of slowing down in certain segments increases dramatically between delay values of 50 and 100ms.

As shown in Figure 9(b) and (c), the span of the tempo difference from baseline values increases with delay between 0 and 50 ms, and peaks and remains high between latencies of 50 to 100 ms. At 150 ms, the variance in the tempo difference across all segments is less than the levels found between 50 and 100 ms.

(a) Segment Tempo Difference vs. Delay



(b) Tempo Difference Span vs. Delay



(c) Magnitude of Tempo Difference Span vs. Delay

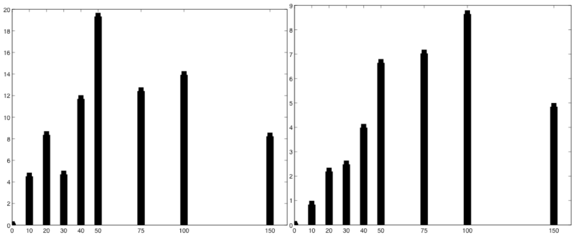


Figure 10. Segmental tempo difference analysis of movement II: Rustique.

5.1.2. Analysis of Movement II: Rustique

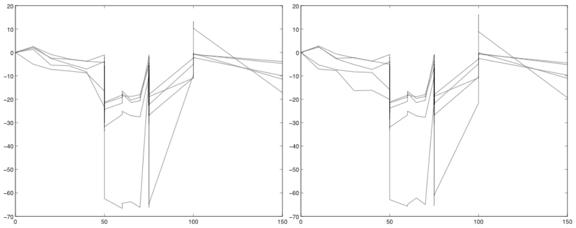
Figure 10 shows the segmental tempo difference analysis for performances of the slower Rustique. In general, the degree of difference from the baseline performance is noticeably reduced, suggesting that the baseline tempo of the piece affects the magnitude of the variance in the tempo difference.

Major slowing down of the short final segment (in measure 22) in the Prima part, and of several segments in the Seconda part, can be observed between 50 and 100 ms in Figure 10(a). Again, the span of the tempo difference across all segments peaks and remains high at latency values between 50 and 100 ms. The tempo difference span is not as large at 150 ms as the values found between 50 and 100 ms.

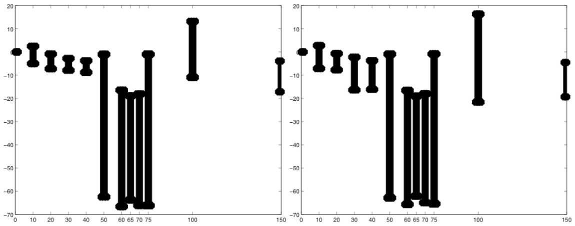
5.1.3. Analysis of Movement III: Final

Figure 11 shows the segmental tempo difference analysis for performances of the rapidly-paced Final, the fastest movement of the three. Relatively little tempo difference is observed for delay values less than 50 ms.

(a) Segment Tempo Difference vs. Delay



(b) Tempo Difference Span vs. Delay



(c) Magnitude of Tempo Difference Span vs. Delay

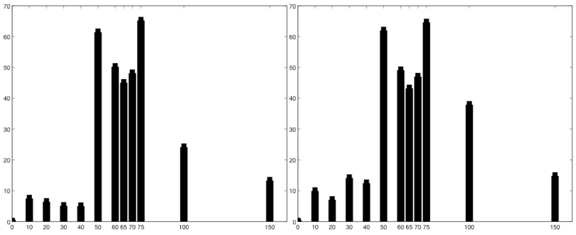


Figure 11. Segmental tempo difference analysis of movement III: Final.

Between 50 and 100 ms, one particular segment is drastically extended (segment 71-72), the final bar of the movement where the two players coordinate an entrance after a period of silence. Even without considering this outlier, the peaks in the tempo difference span occur between 50 and 100 ms. The increased number of performances at delays between 50 and 75ms allow us to improve the density of the data points in this region. Again, the tempo difference span at 150 ms delay is less than those found between 50 and 100 ms delay.

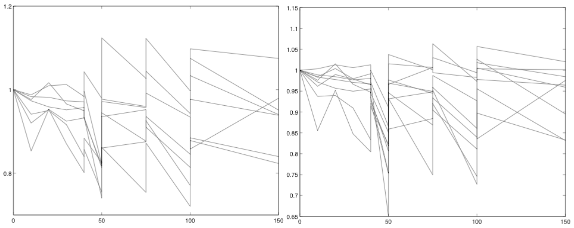
5.1.4. Discussion

We observed that the variability in segmental tempo is markedly increased around the delay thresholds, between 50 and 100 ms for movements 1 and 2, and between 50 and 75 ms for movement 3.

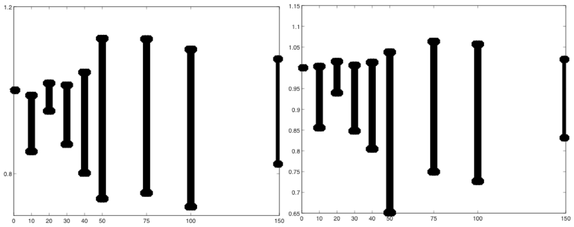
5.2. Segmental Tempo Scaling

The charts in Sections 5.2.1 through 5.2.3 present the segmental tempo scaling results for the movements: Prelude (Figure 12), Rustique (Figure 13), and Final (Figure 14). The three parts of the figure are: (a) graphs of the tempo scaling values over auditory delay, plots of data from all segments are overlaid on the same graph; (b) graphs highlighting the span of the tempo scaling plots; and, (c) plots of the magnitude of the tempo span against auditory delay.

a) Segment Tempo Scaling versus Delay



(b) Tempo Scaling Span vs. Delay



(c) Magnitude of Tempo Scaling Span vs. Delay

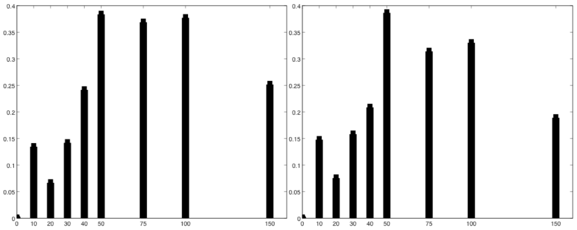


Figure 12. Segmental tempo scaling analysis of movement I: Prelude.

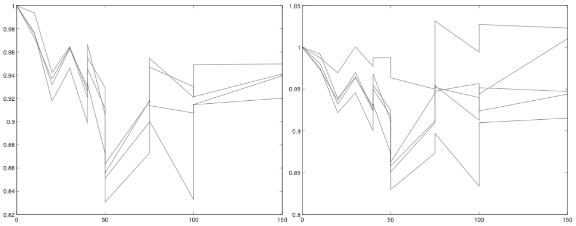
5.2.1. Analysis of Movement I: Prelude

Figure 12 shows the segmental tempo ratio analysis of performances of the Prelude. Similar to the tempo difference analysis, the tempo ratio varies most widely between 50 and 100ms, as shown by all three sets of plots (a), (b) and (c).

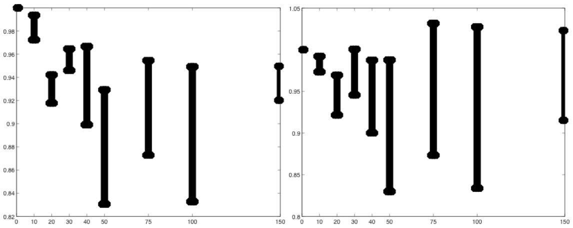
5.2.2. Analysis of Movement II: Rustique

Figure 13 shows the segmental tempo scaling analysis of performances of the Rustique. The tempo ratio varies most widely across all segments between 50 and 100ms, and moderately between 100 and 150ms.

a) Segment Tempo Scaling versus Delay



(b) Tempo Scaling Span vs. Delay



(c) Magnitude of Tempo Scaling Span vs. Delay

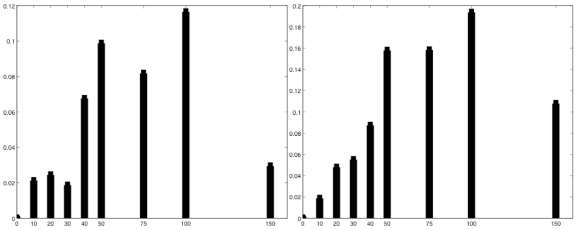


Figure 13. Segmental tempo scaling analysis of movement II: Rustique.

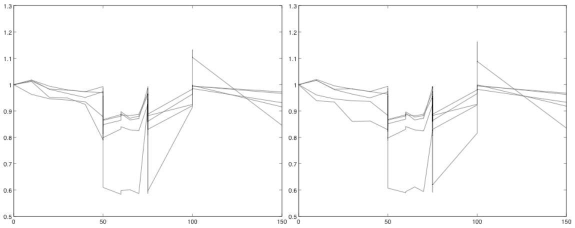
5.2.3. Analysis of Movement III: Final

Figure 14 shows the segmental tempo scaling analysis of performances of the Final. Similar to the tempo difference analyses for this movement, the tempo ratio varies most widely between 50 and 75ms.

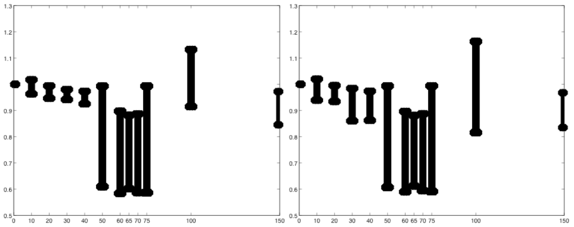
5.3. Summary and Discussion

In summary, both the tempo difference and tempo scaling analyses results showed that there was increased variability in performance strategies (segmental tempo) between auditory delays of 50 to 100 ms for movement 1 and 2, and 50 to 75 ms for movement 3. The tempo difference and tempo scaling climbed steadily between 0 and 50 ms delay, and surprisingly decreased again after 100 ms. Figure 15 summarizes the results of our quantitative analyses.

a) Segment Tempo Scaling versus Delay



(b) Tempo Scaling Span vs. Delay



(c) Magnitude of Tempo Scaling Span vs. Delay

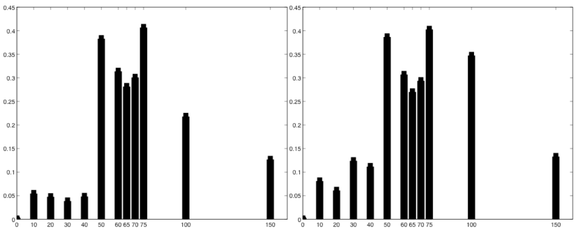


Figure 14. Segmental tempo scaling analysis of movement III: Final.

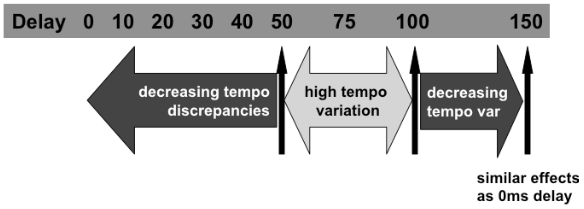
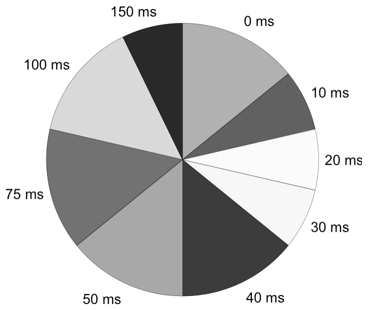


Figure 15. Summary of quantitative analysis of psychophysical experiments.

One might suspect that the variability in the segmental tempo difference and scaling values could be influenced by the number of data points collected at each delay value. Figure 16 shows the number of performances recorded at each delay value for (a) movements 1 and 2, and (b) movement 3.

Figure 16(a) shows a relatively even distribution of performances collected across all delay values. Movement 3 was selected for multiple takes under Experiment C, and hence shows a larger number of data points at 50 and 75 ms. However, the analysis charts for movement three (Figures 11 and 14) show the same degree of increased variability for data points between 50 and 75 ms as those at these boundary values.

(a) Number of performances of mvts 1&2 (total=14 each)



(b) Number of performances of mvt 3 (total=28)

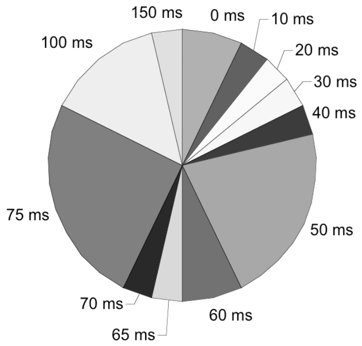


Figure 16. Number of performances at each delay.

Note that the span of the tempo difference and scaling charts is the result of not only the tempo variability across multiple performances, but also of tempo variability across different segments of the same performance.

One surprising observation in our analyses was that tempo variability across segments and performances was reduced at delay value 150ms. A possible explanation for this phenomenon might be that delays of such magnitude are so unacceptable that the players were performing on ‘auto-pilot’, not so much listening and strategizing as simply utilizing their practiced mode of playing. Another explanation could be that unacceptably high delays induce the use of more stable strategies that tend not to vary greatly from the practiced norm.

6. CONCLUSIONS

We have presented a segmental tempo analysis of performances recorded in user experiments for synchronous musical collaboration in a networked environment. The results of these analyses show that methods that quantify performance decisions in recorded data can be a useful tool in studying musician and ensemble behaviour in virtual environments.

From prior analyses of the user responses, we found that, in order to create a collaborative performance, players need to share a common viewpoint to reasonably formulate and work towards common goals. Having this common perspective appears to be more important than having an immediate feedback of the user's own actions. The users' self-reported auditory latency threshold is between 50 and 75 ms.

Quantitative analyses of tempo difference and scaling from a baseline performance showed marked increase in variability between 50 and 100ms for movements 1 and 2, and movement 3. Delays between 0 and 50ms produced tempo variability (both difference and scaling) that increased steadily with delay; delays greater than 100ms produced lower tempo variability.

We hypothesize that greater tempo variability is observed around the usability threshold because around this delay range, the players are exploring new strategies to compensate for the delay. Beyond the threshold, the delay is unacceptable and the players revert to their modus operandi and vary less from the practiced tempi.

Future studies will incorporate more quantitative analyses of, and measures for, musical synchronization and adaptation, and qualitative analyses of the videos and transcripts of the debriefing sessions.

7. ACKNOWLEDGEMENTS

We thank our musical collaborators Vely Stoyanova and Ilia Tosheff (the Tosheff Piano Duo) for their valuable contributions, Christos Papadopoulos for the low-latency audio streaming technology, and Chris Kyriakakis for the use of his Protools system to control auditory delay. We acknowledge the help of numerous student research assistants, who contributed at various times to system development, including Hui-Yun Kao, Dwipal Desai, Kanika Malhotra, Moses Pawar, Rishi Sinha, and Shiva Sundaram. We thank Will Meyer for his camera work and video editing, and Seth Scafani and Allan Weber for providing technical support throughout the project.

Finally, the DIP experiments were made possible by the National Science Foundation grants MRI-0321377, No. 0347988, and Cooperative Agreement No. EEC-9529152. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect those of the National Science Foundation.

8. REFERENCES

[1] Chafe, C., Gurevich, M., Leslie, G., and Tyan, S. "Effect of Time Delay on Ensemble Accuracy." *Proceedings of the International Symposium on Musical Acoustics (ISMA2004),* Nara, Japan, March 31-April 3, 2004.

[2] Chafe, C., Wilson, S., Leistikow, R., Chisholm, D., Scavone, G. "A Simplified Approach to High Quality Music and Sound over IP", *Proceedings of the Workshop on Digital Audio Effects (DAFx-00),* Verona, Italy, pp. 159-163, 2000.

[3] Cain, C., Cooperstock, J. and Woszczyk, W., "Multichannel Audio Over the Internet2: The McGill to USC Demonstration", *109th Audio Engineering Society Convention*, Los Angeles, CA, September 2000.

[4] Chew, E., Sawchuk, A.A., Zimmerman, R., the Tosheff Piano Duo (Stoyanova, V. and Tosheff, I.), Kyriakakis, C., Papadopoulos, C., François, A.R.J., Volk, A. (2004). "Distributed Immersive Performance", *Proceedings of the 2004 Annual National Association of the Schools of Music (NASM) Meeting,* San Diego, CA, November 22, 2004.

[5] Chew, E., Zimmermann, R., Sawchuk, A.A., Kyriakakis, C., Papadopoulos, C., François, A.R.J., Kim, G., Rizzo, A., and Volk, A. "Musical Interaction at a Distance: Distributed Immersive Performance", *Proceedings of the MusicNetwork Fourth Open Workshop on Integration of Music in Multimedia Applications,* Universitat Pompeu Fabra, Barcelona, Spain. Co-located with WEDELMUSIC 2004 Conference, September 15-16, 2004.

[6] Cooperstock, J. McGill University's CIRMMT – *http://cim.mcgill.ca/sre*

[7] Integrated Media Systems Center: *http://imsc.usc.edu*

[8] Gresham-Lancaster, S. AB_Time II Tri-location performance over the Internet in Marseilles, FR, Troy, NY and, Oakland, CA. Vancouver-Marseilles-Troy concert at NIME 2005. January 25 and 26, 2005.

[9] Kapur, A., Wang, G., Davidson, P., Cook, P.R., Trueman, D., Park, T.H., and Bhargava, M. "The Gigapop Ritual: A Live Networked Performance Piece for Two Electronic Dholaks, Digital Spoon, DigitalDoo, 6 String Electric Violin, Rbow, Sitar, Tabla, and Bass Guitar," *New Interface for Musical Expression (NIME)*, Montreal, Canada, May 23, 2003.

[10] Kyriakakis, C. "Virtual Loudspeakers and Virtual Microphones for Multichannel Audio", IEEE International Conference on Consumer Electronics, Los Angeles, June 15th, 2000.

[11] Sawchuk, A.A., Chew, E., Zimmermann, R., Papadopoulos, C., and Kyriakakis, C., "From

Remote Media Immersion to Distributed Immersive Performance", *Proceedings of the ACM SIGMM 2003 Workshop on Experiential Telepresence (ETP 2003)*, Berkeley, California, USA. In conjunction with ACM Multimedia 2003, November 7, 2003.

[12] Schooler, E. M., Touch, J. "Distributed Music - A Foray into Networked Performance", *International Network Music Festival*, Electronic Cafe, Santa Monica, CA Sept. 21, 1993. Musical Distractions Group – *http://www.async.caltech.edu/~schooler/music.html*

[13] Schuett, N. "The Effects of Latency on Ensemble Performance." Undergraduate Honors Thesis, Stanford CCRMA, May 2002.

[14] Shahabi, C., Zimmermann, R., Fu, K., and Yao, S.-Y.D., "Yima: A Second Generation Continuous Media Server", *IEEE Computer*, vol. 35, pp. 56–64, June 2002.

[15] Stanford University's SoundWIRE Group – *http://ccrma.stanford.edu/groups/soundwire*

[16] Tosheff Piano Duo – *http://www.tosheffpianoduo.com*

[17] Zimmermann, R. and Fu, K. "Comprehensive Statistical Admission Control for Streaming Media Servers", *Proceedings of the 11th ACM International Multimedia Conference (ACM Multimedia 2003)*, Berkeley, California, November 2-8, 2003.

[18] Zimmermann, R., Fu, K., and Desai, D.A. "HYDRA: High-performance Data Recording Architecture for Streaming Media", Book chapter in *Video Data Management and Information Retrieval*, editor Sagarmay Deb, University of Southern Queensland, Toowoomba, QLD 4350, Australia. Published by Idea Group Inc., publisher of the Idea Group Publishing, Information Science Publishing and IRM Press imprints, 2004.

[19] Zimmermann, R., Fu, K., and Ku, W.-S. "Design of a Large Scale Data Stream Recorder", *Proceedings of the 5th International Conference on Enterprise Information Systems (ICEIS 2003)*, Angers, France, April 23-26, 2003.

[20] Zimmermann, R., Kyriakakis, C., Shahabi, C., Papadopolous, C., Sawchuk, A.A., and Neumann, U. "The Remote Media Immersion System", *IEEE MultiMedia Magazine*, vol. 11, no. 2, pp. 48-57, April-June 2004.